DEEP UNFOLDING-BASED IMAGE RECONSTRUCTION FOR QUANTA IMAGE SENSORS

Wataru Otobe[†]*, Kosuke Kurihara*[†]*, Yoshihiro Maeda*[‡]*, and Takayuki Hamamoto*[†]

[†]Tokyo University of Science, Japan [‡]Shibaura Institute of Technology, Japan

ABSTRACT

Quanta image sensors are an emerging type of image sensor offering single-photon sensitivity. In this paper, we propose a deep unfolding-based image reconstruction method that integrates an alternating direction method of multipliers optimization with a total variation prior. The proposed method effectively combines model-based approaches with deep learning, providing enhanced interpretability and performance. Experimental results demonstrate that our approach outperforms conventional methods based on the observation model for quanta image sensors.

***Index Terms*—** Quanta image sensor (QIS), deep unfolding, image reconstruction

1. INTRODUCTION

Quanta image sensors (QISs) are an emerging type of image sensor offering single-photon sensitivity [1]. Unlike conventional image sensors, such as CCD and CMOS sensors, which output the electron charge accumulated from incoming photons during the exposure time, QISs can output the photon detection status in digital format on a photon-by-photon basis at ultra-high speed. This unique capability provides two key advantages to QIS-based imaging systems: extremely high light sensitivity and ultra-high temporal resolution. Therefore, QIS-based imaging has potential applications, including high dynamic range imaging [2], [3], [4], low-light environment imaging [3], [4], [5], and motion estimation for high-speed moving objects [3], [4], [6], [7].

QIS-based imaging systems comprise two primary steps: photon detection and image reconstruction. During the photon detection step, the system determines incident photons for each pixel at extremely high speed and then produces multiple binary frames, called *bit-plane images*. In the subsequent image reconstruction step, the system estimates light intensity from these bit-plane images to obtain a multi-bit image. Since the input for the reconstruction process consists of a 1-bit time series with inherently limited information, several reconstruction methods have been developed to obtain high-accuracy images [8], [9], [10].

The conventional image reconstruction methods based on physical observation models [8], [11] have been proposed for reconstructing multi-bit images from bit-plane images. These models utilize the statistical behavior of incident photons because incident photons follow a Poisson process [11]. Using this model, they estimate the number of incident photons and reconstruct the multi-bit image through maximum likelihood estimation (MLE).

To improve image reconstruction performance, several researchers have incorporated prior knowledge of natural images into the statistical models [9], [10]. We refer to such approaches simply as model-based methods. In [9], the authors introduced prior knowledge of spatial characteristics into the statistical model by employing the total variation (TV) norm. They then solved the corresponding optimization problem using the alternating direction method of multipliers (ADMM). Chan *et al.* [10] extended this approach by integrating existing image denoising algorithms into the ADMM optimization framework, a strategy known as the Plug-and-Play (PnP) algorithm [12]. By incorporating high-performance denoisers such as BM3D and deep neural network (DNN)-based methods, the method can significantly improve the quality of reconstructed images. However, these model-based methods have a notable drawback: their reconstruction performance highly depends on the values of parameters, such as the regularization weight and step size.

In contrast to model-based methods, DNN-based methods have been proposed [5], [13]. Their image reconstruction performance often outperformed that of traditional model-based methods. However, to achieve high performance, these methods require a large number of parameters and a large dataset. Moreover, the DNN-based methods often discard domain knowledge, such as physical observation models and prior knowledge of natural images, and thus lack interpretability and reliability.

To bridge the gap between model-based and DNN-based

This work was supported by JSPS KAKENHI Grant Numbers JP24K02964.

This article has been accepted for publication in IEEE International Conference on Image Processing (ICIP) 2025. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ICIP55913.2025.11084297

methods, researchers have developed *deep unfolding* [14], [15]. The deep unfolding approach transforms iterative optimization algorithms into trainable deep networks [14], [15], combining the strengths of model-based and DNN-based methods. This hybrid approach allows the system to learn optimal parameters, such as step sizes and regularization weights, leading to better convergence and improved reconstruction quality. Moreover, unlike conventional DNN-based methods, deep unfolding preserves the interpretability of the underlying physical model.

In the context of image reconstruction for QIS, only one study applying the deep unfolding approach has been reported [16]. Remez *et al.* [16] proposed a patch-based image reconstruction method based on a sparse coding framework. By incorporating the obtained sparse representation as an image prior and the QIS imaging model, they constructed an image reconstruction method based on the iterative shrinkage thresholding algorithm (ISTA) using a deep unfolding approach [17]. This is a pioneering study introducing deep unfolding for the QIS image reconstruction task. However, the method reconstructs images based on a sparse representation of local regions, leading to degradation when averaging patches to reconstruct the full-size image. Additionally, it requires the separate construction of a sparse dictionary of natural images and an iterative algorithm for image reconstruction; thus, it is a complex formulation and not a straightforward approach.

In this paper, we propose a deep unfolding-based image reconstruction method for QIS-based systems. The proposed method solves the statistical observation model with a TV norm as a natural image prior using a deep unfolding-based ADMM framework. Our approach is simple and straightforward, unlike the previous method [16]. Since it is based on the statistical observation model-based framework for QIS image reconstruction, we can effectively utilize established domain knowledge from image processing research. Our method employs the deep unfolding approach, effectively integrating the advantages of model-based and DNN-based approaches to achieve highly accurate and interpretable image estimation.

The main contributions of this paper are as follows: 1) We formulate a deep unfolding-based image reconstruction method for QIS imaging, incorporating the TV norm as a natural image prior. This method follows the ADMM framework [9], which allows the integration of advancements from both model-based and DNN-based approaches. 2) The proposed method introduces a simple and efficient deep unfolding approach, unlike the conventional deep unfolding-based method [16]. In addition, the proposed method achieves higher performance with fewer parameters than existing approaches.

2. QIS IMAGING MODEL

2.1. Statistical Modeling of Incident Photons

QIS can capture incoming light in photon units, which follows a Poisson process [11]. The probability of y_n photons entering at $n = (1, \cdots, N)$-th pixel can be modeled as

$$p(y_n \,|\, x_n) = \frac{{x_n}^{y_n} e^{-x_n}}{y_n!}, \tag{1}$$

where x_n denotes the light intensity at the surface of the n-th pixel.

When the photon detection threshold of QIS is set to a single photon, the photon detection probability of QIS is determined as

$$p(b_n \,|\, x_n) = \begin{cases} e^{-x_n} & \text{if } \; b_n = 0 \\ 1 - e^{-x_n} & \text{if } \; b_n = 1 \,, \end{cases} \tag{2}$$

where $b_n \in \{0, 1\}$ denotes the binary output of the n-th pixel of QIS; zero and one represent non-detection and detection status, respectively.

2.2. MLE based Image Reconstruction

The objective of image reconstruction is to estimate the light intensity, representing the image, $\boldsymbol{x} = [x_1, ..., x_N]$, from the T binary output frames, $\boldsymbol{b}_{1:T}$, which are called bit-plane images. Because the bit-plane image is binary and contains limited amplitude information, multiple bit-plane images are required to reconstruct a multi-bit image accurately.

By incorporating the statistical model of the incident photon into the maximum likelihood estimation (MLE) framework, the reconstructed image $\boldsymbol{x}$ can be obtained. Specifically, the MLE solution $\hat{\boldsymbol{x}}^{\,\mathrm{MLE}}$ can be estimated by minimizing the following negative log-likelihood as

$$\hat{\boldsymbol{x}}^{\,\mathrm{MLE}} = \arg\min_{\boldsymbol{x}} - \sum_{t=1}^{T} \ln(\, p(\boldsymbol{b}_t \,|\, \boldsymbol{x})\,), \tag{3}$$

where $\boldsymbol{b}_t = [b_{1,t}, ..., b_{N,t}]$ denotes the t-th bit-plane image.

Based on the statistical modeling of incident photons (Eq. (2)), the cost function, $f(\boldsymbol{x}) \equiv -\sum_{t=1}^{T} \ln(\, p(\boldsymbol{b}_t \,|\, \boldsymbol{x})\,)$, is expressed as follows:

$$f(\boldsymbol{x}) = \sum_{n=1}^{N} \left\{ -S_n^{(0)} \ln\left(e^{-\frac{x_n}{T}}\right) - S_n^{(1)} \ln\left(1 - e^{-\frac{x_n}{T}}\right) \right\}, \tag{4}$$

where $S_n^{(1)} = \sum_{t=1}^{T} b_{n,t}$ and $S_n^{(0)} = T - S_n^{(1)}$ denote the total number of $b_{n,t}$ that are ones and zeros, respectively.

The MLE solution can be obtained in a closed-form expression. The n-th pixel of $\hat{\boldsymbol{x}}^{\,\mathrm{MLE}}$, denoted by $\hat{x}_n^{\,\mathrm{MLE}}$, is represented as

$$\hat{x}_n^{\,\mathrm{MLE}} = -\ln\left(1 - \frac{S_n^{(1)}}{T}\right). \tag{5}$$

3. PROPOSED METHOD

3.1. Overview

Figure 1 shows an overview of the proposed method. We construct a deep unfolding network by unfolding an iterative model-based image reconstruction algorithm. To this end, we design an iterative image reconstruction algorithm based on maximum a posteriori (MAP) estimation that incorporates prior knowledge for natural images and photon statistical modeling. We then unfold the iterative algorithm into the deep network with trainable parameters.

3.2. Problem Formulation Based on MAP Estimation

We first formulate the image reconstruction problem based on the MAP estimation framework [9]. The latent image $\boldsymbol{x}$ is obtained by minimizing the following negative log-posterior function as

$$\begin{aligned}\hat{\boldsymbol{x}} &= \arg\min_{\boldsymbol{x}} -\ln p(\boldsymbol{x} \mid \boldsymbol{b}_{1:T}) \\ &= \arg\min_{\boldsymbol{x}} \underbrace{-\sum_{t=1}^{T} \ln(\, p(\boldsymbol{b}_t \mid \boldsymbol{x})\,)}_{f(\boldsymbol{x})} \underbrace{-\ln(\, p(\boldsymbol{x})\,)}_{g(\boldsymbol{x})}, \end{aligned} \tag{6}$$

where $p(\boldsymbol{x} \mid \boldsymbol{b}_{1:T})$ and $\boldsymbol{b}_{1:T} = (\boldsymbol{b}_1, ..., \boldsymbol{b}_T)$ denote prior probability and the set of bit-plane images, respectively.

We explain the operation of each term on the right-hand side in Eq. (6). The first term, denoted by $f(\boldsymbol{x})$, evaluates the consistency of $\boldsymbol{x}$ with the incident photon model, which is the same as in MLE (Sect. 2.2).

The second term, denoted by $g(\boldsymbol{x})$, operates as a regularization in which we can model based on prior knowledge on natural image statistics, such as sparsity and self-similarity [18]. In our method, we employ total variation (TV), which models the sparsity in the image gradient [19]. Our TV-based regularization term $g(\boldsymbol{x})$ is represented as

$$g(\boldsymbol{x}) \equiv \lambda \left(\|\boldsymbol{D}_h\, \boldsymbol{x}\|_1 + \|\boldsymbol{D}_v\, \boldsymbol{x}\|_1 \right), \tag{7}$$

where λ and $\|\cdot\|_1$ denote the weight parameter and l_1 norm, respectively. In addition, $\boldsymbol{D}_h$ and $\boldsymbol{D}_v$ denote the matrices calculating the horizontal and vertical differences of the input image, respectively.

3.3. Iterative Optimization Algorithm

To solve the cost function presented in the previous section, we introduce an alternative direction method of multipliers (ADMM) framework [20]. To apply ADMM, we first rewrite the cost function (Eq. (6)) by using the auxiliary variable $\boldsymbol{z}$ as

$$f(\boldsymbol{x}) + \lambda \|\boldsymbol{z}\|_1 \quad \text{s.t. } \boldsymbol{z} = \boldsymbol{D}\boldsymbol{x}\,, \tag{8}$$

where $\boldsymbol{D} = [\boldsymbol{D}_h^\top, \boldsymbol{D}_v^\top]^\top$ represents the stacked difference matrix.

The augmented Lagrangian function of Eq. (8) is represented as

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{v}) = f(\boldsymbol{x}) + \lambda \|\boldsymbol{z}\|_1 + \boldsymbol{v}^\top (\boldsymbol{D}\boldsymbol{x} - \boldsymbol{z}) + \frac{\rho}{2} \|\boldsymbol{D}\boldsymbol{x} - \boldsymbol{z}\|_2^2\,, \tag{9}$$

where $\boldsymbol{v}$ and ρ denote the Lagrange multiplier and the penalty parameter, respectively. In addition, $\|\cdot\|_2$ denotes the l_2 norm. The ADMM algorithm alternately updates variables $(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{v})$ in an iterative manner. Eq. (9) can be solved by iteratively solving the following sub-problems:

$$\boldsymbol{x}^{(k)} = \arg\min_{\boldsymbol{x}} \left\{ f(\boldsymbol{x}) + \frac{\rho}{2} \left\| \boldsymbol{D}\boldsymbol{x} - \boldsymbol{z}^{(k-1)} + \frac{\boldsymbol{v}^{(k-1)}}{\rho} \right\|_2^2 \right\}, \tag{10}$$

$$\boldsymbol{z}^{(k)} = \mathrm{prox}_{\rho/\lambda,\, \|\cdot\|_1} \left(\boldsymbol{D}\boldsymbol{x}^{(k)} + \frac{\boldsymbol{v}^{(k-1)}}{\rho} \right), \tag{11}$$

$$\boldsymbol{v}^{(k)} = \boldsymbol{v}^{(k-1)} + \rho \left(\boldsymbol{D}\boldsymbol{x}^{(k)} - \boldsymbol{z}^{(k)} \right), \tag{12}$$

where k denotes the iteration index. In addition, $\mathrm{prox}_{\beta,\, \|\cdot\|_1}(\boldsymbol{u})$ represents the proximal operator and is defined as

$$\mathrm{prox}_{\beta,\, \|\cdot\|_1}(\boldsymbol{u}) \equiv \arg\min_{\boldsymbol{w}} \left\{ \|\boldsymbol{w}\|_1 + \frac{\beta}{2} \|\boldsymbol{w} - \boldsymbol{u}\|_2^2 \right\}. \tag{13}$$

We describe how to solve Eqs. (10) and (11). For Eq. (10), it can be solved by using a gradient descent method because Eq. (10) is differentiable with respect to $\boldsymbol{x}$. The latent image $\boldsymbol{x}^{(k)}$ at l-th inner iteration, denoted as $\boldsymbol{x}^{(k,l)}$, is updated iteratively using the steepest gradient descent method:

$$\begin{aligned}&\boldsymbol{x}^{(k,l)} \leftarrow \boldsymbol{x}^{(k,l-1)} \\ &- \eta \left\{ \nabla f(\boldsymbol{x}^{(k,l-1)}) + \rho \boldsymbol{D}^\top \left(\boldsymbol{D}\boldsymbol{x}^{(k,l-1)} - \boldsymbol{z}^{(k)} + \frac{\boldsymbol{v}^{(k)}}{\rho} \right) \right\}, \end{aligned} \tag{14}$$

where η and ∇ are a step size parameter and a gradient operator, respectively.

For Eq. (11), we utilize the proximal gradient approach because it involves the non-differentiable l_1 norm, as shown in Eq.(13). Based on the proximal algorithm, Eq. (13) can be evaluated by applying element-wise soft thresholding [20]. The calculation for the j-th element of $\boldsymbol{u}$ in Eq. (13), denoted by u_j, is represented as

$$\mathrm{prox}_{\beta,\, \|\cdot\|_1}(u_j) = \begin{cases} u_j + \beta & (u_j < -\beta) \\ 0 & (-\beta \le u_j \le \beta) \\ u_j - \beta & (\beta < u_j)\,. \end{cases} \tag{15}$$

3.4. Network Construction via Deep Unfolding

We unfold the above-mentioned iterative algorithm into a deep network based on a deep unfolding framework. As

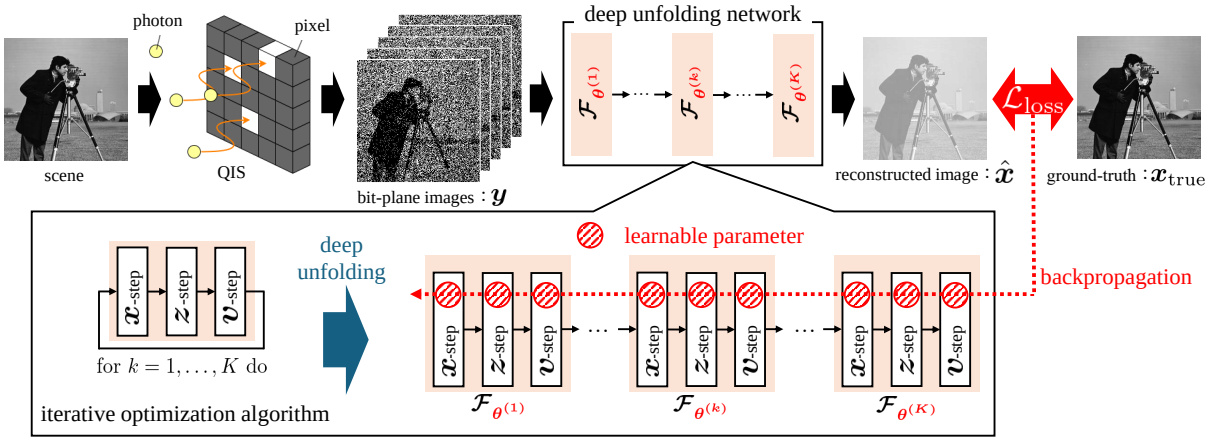


Fig. 1: Overview of the proposed method. Based on QIS-imaging model and prior knowledge on natural images, we construct deep unfolding network for QIS image reconstruction task.

shown in Figure 1, we transform the set of sub-problems for each iteration (Eqs. (10) to (12)) into a single layer of a deep network. The input and output for the k-th layer of our network are represented as

$$\boldsymbol{x}^{(k)}, \boldsymbol{z}^{(k)}, \boldsymbol{v}^{(k)} \leftarrow \mathcal{F}_{\theta^{(k)}}\left(\boldsymbol{x}^{(k-1)}, \boldsymbol{z}^{(k-1)}, \boldsymbol{v}^{(k-1)}\right), \quad (16)$$

where $\mathcal{F}_{\theta^{(k)}}(\cdot)$ is the operator corresponding to the k-th layer, composed of Eqs. (10) to (12), with a trainable parameter $\theta^{(k)}$. The parameter of the k-th layer $\theta^{(k)}$ is composed as

$$\theta^{(k)} = \left(\lambda^{(k)},\ \rho^{(k)},\ \{\eta^{(k,l)}\}_{l=1}^{L}\right), \quad (17)$$

where L is the number of inner steps originated from Eq. (14).

We train the set of parameters $\{\theta^{(k)}\}_{k=1}^{K}$ through the back-propagation procedure. We minimize the following mean-square-error (MSE) loss function in a supervised manner as

$$\mathcal{L}_{loss} = \|\hat{\boldsymbol{x}} - \boldsymbol{x}_{\text{true}}\|_2^2, \quad (18)$$

where $\hat{\boldsymbol{x}}$ and $\boldsymbol{x}_{\text{true}}$ are the reconstructed image through our deep unfolding network and ground-truth image, respectively.

4. EXPERIMENTS

4.1. Settings

To demonstrate the effectiveness of the proposed method, we conducted experiments using simulated bit-plane images. Our experiments were run on Ubuntu with an Intel Core i7-14700KF and GeForce RTX 4090.

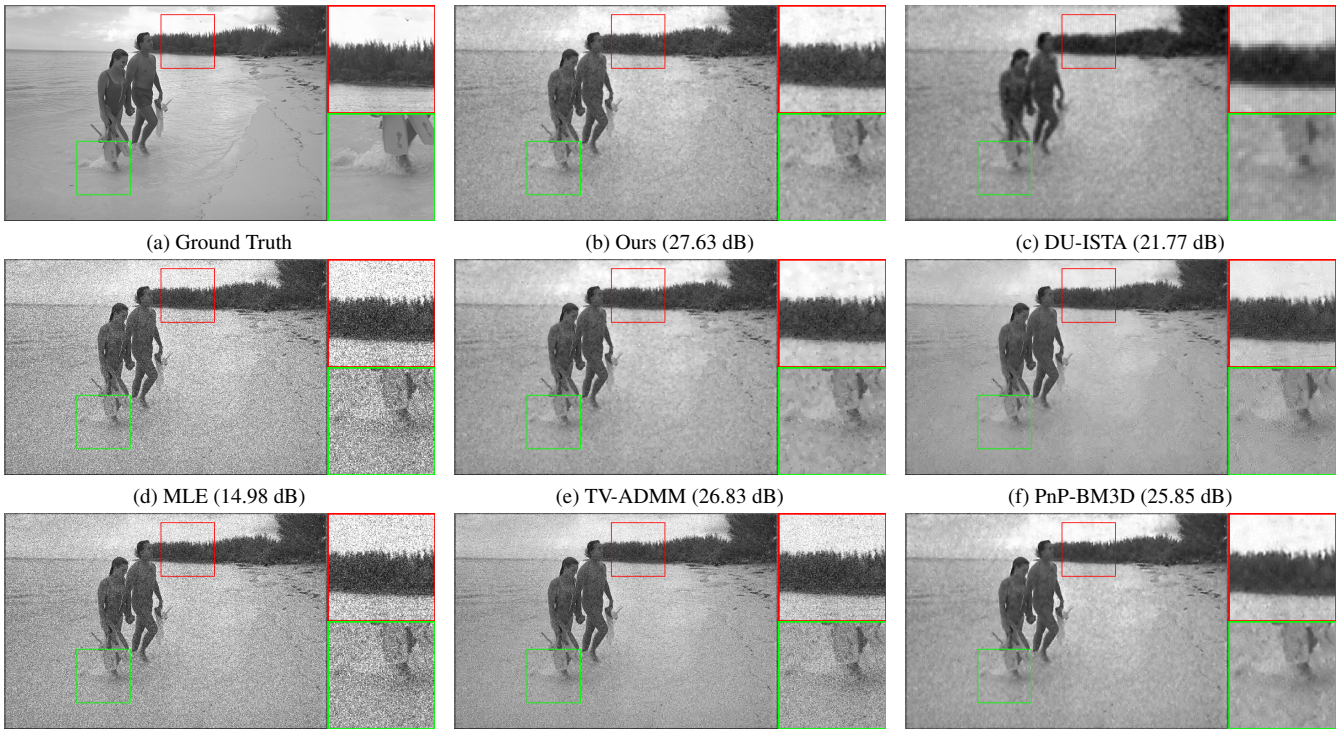
Datasets: As a training dataset, we used 2060 grayscaled images from the UPenn dataset [21]. We randomly cropped the original images to 256×256 pixels and used these images as ground truth. Similarly to previous works [5], [7], we simulated T bit-plane images from a ground truth scene. We varied $T = 8, 16, 32$ to analyze the influence of the number of input bit-plane images T. For the test dataset, we used 24 images from the Kodak dataset [22]. We simulated bit-plane images in the same manner as in the training phase. We used these simulated bit-plane images as input for our method and the comparison methods.

Network Training: We set the number of layers K and inner layers L as 50 and 1, respectively. We initialize the learnable parameters of our network for each layer as $\lambda^{(k)} = 0.10$, $\rho^{(k)} = 0.30$, and $\eta^{(k,l)} = 0.10$ ($\forall k \in \{1, \cdots, K\}$, $\forall l \in \{1, \cdots, L\}$). We used the AdamW optimizer with a learning rate of 0.0001 and a batch size of 64. In addition, the number of epochs was set to 100.

Comparison Methods: We compared the proposed method with seven methods, including deep unfolding-based, model-based, PnP-based, and DNN-based methods. For deep unfolding-based methods, we included DU-ISTA [16], which adopts a learned ISTA. For model-based methods, we used MLE [8] and TV-ADMM [9]. MLE applies MLE in a straightforward manner. TV-ADMM employs the ADMM framework with TV regularization, which corresponds to the proposed method without deep unfolding. For PnP-based methods, we used PnP-BM3D [10], PnP-DnCNN [23], and PnP-DRUNet [24]. For deep learning-based methods, we also compared with QISNet [13], a fully trainable DNN-based approach.

4.2. Results

Table 1 presents the quantitative comparison results using PSNR. The proposed method achieves the best performance at $T = 32$ and the second-best performance at $T = 8, 16$. For $T = 8, 16$, the proposed method provides comparable performance to QISNet, a DNN-based approach. Although DU-ISTA follows a similar deep unfolding framework, the proposed method achieves higher PSNR, demonstrating its



(a) Ground Truth (b) Ours (27.63 dB) (c) DU-ISTA (21.77 dB)

(d) MLE (14.98 dB) (e) TV-ADMM (26.83 dB) (f) PnP-BM3D (25.85 dB)

(g) PnP-DnCNN (16.74 dB) (h) PnP-DRUNet (20.55 dB) (i) QISNet (26.94 dB)

Fig. 2: Visual comparison of reconstructed images at $T = 32$.

Table 1: Quantitative comparison of average PSNR [dB] in Kodak dataset. Best and second best scores are in **bold** and underline.

Method	$T = 8$	$T = 16$	$T = 32$
Ours	<u>22.75</u>	<u>24.44</u>	**25.99**
DU-ISTA [16]	19.78	18.86	20.47
MLE [8]	10.28	13.71	17.07
TV-ADMM [9]	16.25	20.80	25.00
PnP-BM3D [10]	16.31	21.83	25.53
PnP-DnCNN [23]	12.66	15.28	18.06
PnP-DRUNet [24]	13.77	17.20	21.20
QISNet [13]	**23.65**	**24.67**	<u>25.70</u>

effectiveness. These results confirm that the proposed method exhibits performance that is nearly equivalent to that of the DNN-based approach. These results indicate that the proposed method demonstrates performance that is nearly comparable to the DNN-based approach.

For PnP-based methods, PSNR decreases as T becomes smaller. This decrease occurs because the noise level in the input bit-plane images exceeds the assumptions of the denoisers, which limits their effectiveness. Therefore, it suggests that the plug-and-play incorporation of conventional denoising methods has inherent constraints in improving image reconstruction performance. In contrast, the proposed method integrates the observation model and prior knowledge into a deep network architecture. The end-to-end training manner allows the network to adapt to different noise conditions and learn optimal reconstruction strategies, which leads to improved image reconstruction quality.

Table 2: The comparisons of trainable parameters.

Method	architecture	# of param.
Ours	unfolding	150
DU-ISTA [16]	unfolding	16,448
QISNet [13]	DNN	1,035,201

Figure 2 depicts an example of reconstructed images. It can be seen that the proposed method can effectively reduce noise while preserving detailed structural information.

Table 2 presents a comparison of trainable parameters for deep unfolding and DNN-based methods, both of which rely on trainable parameters. The proposed method requires the fewest trainable parameters among the methods, demonstrating its computational efficiency. Although QISNet achieves the highest image reconstruction performance in most cases, as shown in Table 1, it uses significantly more parameters than the proposed method. These results confirm that the proposed method has both high image reconstruction performance and computational efficiency.

5. CONCLUSION

We proposed the deep unfolding-based image reconstruction method for QIS. Because our deep unfolding network was constructed based on a model-based iterative image reconstruction algorithm, we can effectively utilize domain knowledge, including the QIS-imaging model and the prior knowledge of natural images. Moreover, a deep unfolding framework enables us to train the constructed deep network in a supervised manner. The proposed method has both high image reconstruction performance and computational efficiency compared with conventional model-based, deep unfolding-based, and DNN-based methods.

6. REFERENCES

[1] E. R. Fossum, J. Ma, S. Masoodian, L. Anzagira, and R. Zizza, "The quanta image sensor: Every photon counts," *Sensors*, vol. 16, no. 8, 2016.

[2] J. Ma, D. Zhang, D. Robledo, L. Anzagira, and S. Masoodian, "Ultra-high-resolution quanta image sensor with reliable photon-number-resolving and high dynamic range capabilities," *Scientific Reports*, vol. 12, 2022.

[3] T. Zhang, M. Dutson, V. Boominathan, M. Gupta, and A. Veeraraghavan, " Streaming Quanta Sensors for Online, High-Performance Imaging and Vision," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 47, no. 03, pp. 1564–1577, Mar. 2025.

[4] S. H. Chan, "What does a one-bit quanta image sensor offer?" *IEEE Transactions on Computational Imaging*, vol. 8, pp. 770–783, 2022.

[5] Y. Chi, A. Gnanasambandam, V. Koltun, and S. H. Chan, "Dynamic low-light imaging with quanta image sensors," in *Proc. European Conference on Computer Vision (ECCV)*, 2020, pp. 122–138.

[6] H. Fukawa, K. Kurihara, Y. Maeda, S. Sato, and T. Hamamoto, "Motion estimation for quanta image sensors using spatio-temporal priors," in *Proc. IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 2024, pp. 1–5.

[7] K. Iwabuchi, Y. Kameda, and T. Hamamoto, "Image quality improvements based on motion-based deblurring for single-photon imaging," *IEEE Access*, vol. 9, pp. 30 080–30 094, 2021.

[8] S. H. Chan, O. A. Elgendy, and X. Wang, "Images from bits: Non-iterative image reconstruction for quanta image sensors," *Sensors*, vol. 16, no. 11, 2016.

[9] S. H. Chan and Y. M. Lu, "Efficient image reconstruction for gigapixel quantum image sensors," in *Proc. IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2014, pp. 312–316.

[10] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play admm for image restoration: Fixed-point convergence and applications," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 84–98, 2017.

[11] F. Yang, Y. M. Lu, L. Sbaiz, and M. Vetterli, "Bits from photons: Oversampled image acquisition using binary poisson statistics," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1421–1436, 2012.

[12] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2013, pp. 945–948.

[13] J. H. Choi, O. A. Elgendy, and S. H. Chan, "Image reconstruction for quanta image sensors using deep neural networks," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 6543–6547.

[14] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, 2021.

[15] N. Shlezinger, J. Whang, Y. C. Eldar, and A. G. Dimakis, "Model-based deep learning," *Proceedings of the IEEE*, vol. 111, no. 5, pp. 465–499, 2023.

[16] T. Remez, O. Litany, and A. Bronstein, "A picture is worth a billion bits: Real-time image reconstruction from dense binary threshold pixels," in *Proc. IEEE International Conference on Computational Photography (ICCP)*, 2016, pp. 1–9.

[17] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. International Conference on Machine Learning (ICML)*, 2010, pp. 399–406.

[18] A. Hyvärinen, J. Hurri, and P. O. Hoyer, *Natural Image Statistics*. London, UK: Springer, 2009, vol. 39.

[19] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[20] N. Komodakis and J.-C. Pesquet, "Playing with duality: An overview of recent primal dual approaches for solving large-scale optimization problems," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 31–54, 2015.

[21] G. Tkačik et al., "Natural images from the birthplace of the human eye," *PLOS ONE*, vol. 6, no. 6, pp. 1–12, Jun. 2011.

[22] Eastman Kodak Company, *Kodak lossless true color image suite*, `http://r0k.us/graphics/kodak/`, Accessed: 2025-01-23.

[23] E. K. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin, "Plug-and-play methods provably converge with properly trained denoisers," in *Proc. International Conference on Machine Learning (ICML)*, 2019, pp. 5546–5557.

[24] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 6360–6376, 2022.